



Music expertise shapes audiovisual temporal integration windows for speech, sinewave speech, and music

Hweeling Lee^{1,2*} and Uta Noppeney^{1,3}

¹ Cognitive Neuroimaging Group, Max Planck Institute for Biological Cybernetics, Tübingen, Germany

² Memory Dysfunction in Neurodegenerative Diseases, German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany

³ Computational Neuroscience and Cognitive Robotics Centre, School of Psychology, University of Birmingham, Birmingham, UK

Edited by:

Kaisa Tiippana, University of Helsinki, Finland

Reviewed by:

Mireille Besson, CNRS, France
Argiro Vatakis, Cognitive Systems Research Institute, Athens, Greece

*Correspondence:

Hweeling Lee, Memory Dysfunction in Neurodegenerative Diseases, German Center for Neurodegenerative Diseases (DZNE), Ernst-Robert-Curtius Strasse 12, 53117 Bonn, Germany
e-mail: hweeling.lee@dzne.de

This psychophysics study used musicians as a model to investigate whether musical expertise shapes the temporal integration window for audiovisual speech, sinewave speech, or music. Musicians and non-musicians judged the audiovisual synchrony of speech, sinewave analogs of speech, and music stimuli at 13 audiovisual stimulus onset asynchronies (± 360 , ± 300 , ± 240 , ± 180 , ± 120 , ± 60 , and 0 ms). Further, we manipulated the duration of the stimuli by presenting sentences/melodies or syllables/tones. Critically, musicians relative to non-musicians exhibited significantly narrower temporal integration windows for both music and sinewave speech. Further, the temporal integration window for music decreased with the amount of music practice, but not with age of acquisition. In other words, the more musicians practiced piano in the past 3 years, the more sensitive they became to the temporal misalignment of visual and auditory signals. Collectively, our findings demonstrate that music practicing fine-tunes the audiovisual temporal integration window to various extents depending on the stimulus class. While the effect of piano practicing was most pronounced for music, it also generalized to other stimulus classes such as sinewave speech and to a marginally significant degree to natural speech.

Keywords: multisensory, temporal synchrony, audiovisual integration, plasticity, speech, music

INTRODUCTION

Music training provides a rich multisensory experience that requires integrating signals from different sensory modalities with motor responses. Thus, the musician's brain provides an ideal model to study experience-dependent plasticity in humans (Munte, 2002; Zatorre et al., 2007). Previous research has shown that musicians develop an enhanced auditory system, both at the structural and functional levels (Schlaug et al., 1995; Munte et al., 2002; Schneider et al., 2005; Hannon and Trainor, 2007; Baumann et al., 2008; Imfeld et al., 2009) that seems to benefit linguistic and non-linguistic skills (Magne et al., 2006; Marques et al., 2007; Moreno et al., 2009; Tzounopoulos and Kraus, 2009; Kraus and Chandrasekaran, 2010). Specifically, musicians proved to be better than non-musicians at segmenting speech from background noise (Parbery-Clark et al., 2013), pitch (Besson et al., 2007), and prosodic tasks (Thompson et al., 2004).

Since practicing a musical instrument for an extensive period of time involves precise timing of several hierarchically organized actions, musical expertise may in particular influence the temporal binding of signals across the senses during perception. Even though sensory signals do not have to be precisely synchronous, they have to co-occur within a certain temporal integration window in order to be integrated into a unified percept (Stein et al., 1993; Spence and Squire, 2003; Noesselt et al., 2007, 2008; Lewis and Noppeney, 2010; Stevenson et al., 2011). Recent studies have shown that the temporal integration window can be narrowed or shifted via long-term musical training (Petrini et al., 2009),

short-term perceptual learning (Powers et al., 2009), or short-term audiovisual exposure (Fujisaki et al., 2004). Conversely, it can be widened by exposure to asynchronous stimuli (Navarra et al., 2005).

One critical question is to which extent the impact of musical expertise on audiovisual synchrony perception is specific to the practiced music or whether it generalizes to other stimulus domains. In support of more generic effects, previous studies on auditory processing demonstrated earlier, larger and more robust brainstem responses for musicians relative to non-musicians for both speech and music stimuli (Musacchia et al., 2008; Bidelman and Krishnan, 2010; Bidelman et al., 2011). Moreover, viewing the corresponding videos of the musical instrument in action or facial movements enhanced the temporal and frequency encoding in musicians (Musacchia et al., 2007). Collectively, these results suggest that musical expertise may improve audiovisual processing in a generic fashion at very early processing stages in the brainstem. Based on these results, we may expect that musical expertise fine-tune the temporal integration window generically across multiple stimulus classes such as speech and music.

By contrast, a recent combined psychophysics-fMRI study demonstrated that musicians relative to non-musicians have a significantly narrower temporal integration window for music but not for speech stimuli (Lee and Noppeney, 2011a). Moreover, at the neural level, musicians showed increased audiovisual asynchrony responses and effective connectivity selectively for music but not for speech in a circuitry including the superior temporal

sulcus, the premotor cortex and the cerebellum. These results suggest that music practicing may mold audiovisual temporal binding not only via generic mechanisms of perceptual learning but also via more stimulus-specific mechanisms of sensory-motor learning. More specifically, piano music practicing may fine-tune an internal forward model mapping from action plans specific for piano playing onto visible finger movements and sounds. As this internal forward model furnishes more precise estimates of the relative audiovisual timings of music actions, it sensitizes musicians specifically to audiovisual temporal misalignments of music stimuli. Yet, one may argue that natural speech is not an ideal stimulus class to test whether music expertise transfers from music to other stimulus classes, because both musicians and non-musicians are “speech experts” thereby minimizing any additional effects of musical expertise on audiovisual temporal synchrony perception.

To further investigate whether musical expertise shapes temporal binding of non-music stimuli, we presented 21 musicians and 20 non-musicians participants with natural speech, intelligible sinewave analogs of speech, and piano music stimuli at 13 audiovisual stimulus onset asynchronies (± 360 , ± 300 , ± 240 , ± 180 , ± 120 , ± 60 , and 0 ms) (Dixon and Spitz, 1980; Alais and Burr, 2003; Grant et al., 2004; Zampini et al., 2005; Vatakis and Spence, 2006a,b, 2007, 2008a,b; van Wassenhove et al., 2007; Love et al., 2013). On each trial, participants judged the audiovisual synchrony of natural speech, sinewave speech, and piano music stimuli. We have included these three classes of stimuli to elucidate the main factors that determine whether musical expertise generalizes to other classes of stimuli: Natural speech/sinewave speech and piano music are linked to different motor effectors (mouth vs. hand) and thereby rely on different sensori-motor transformations. By contrast, natural speech and intelligible sinewave speech are identical in the visual facial movements and linguistic representations, but differ in their spectrotemporal structure of the auditory input (Remez et al., 1981; Lee and Noppeney, 2011b; Vroomen and Stekelenburg, 2011; Stekelenburg and Vroomen, 2012; Baart et al., 2014). As sinewave speech is generated by replacing the main speech formants with sinewave analogs, sinewave speech obtains a more musical character. Critically, neither musicians nor non-musicians have been exposed to sinewave speech in their natural environment, so that neither of them are sinewave speech experts. Hence, as with other speech transformations such as rotated speech, both groups should have less precise temporal predictions, and hence, yield a wider temporal integration window for sinewave speech than for piano music or natural speech stimuli (see Maier et al., 2011). These aspects render sinewave speech an ideal stimulus to test for transfer effects from music to other stimulus classes.

Finally, previous studies have demonstrated that humans accumulate statistical information over time for deciding whether auditory and visual signals are synchronous or asynchronous (see Vatakis and Spence, 2006a; Maier et al., 2011). We therefore investigated whether the effect of musical expertise on audiovisual synchrony judgments depends on the stimulus duration by presenting participants with short (piano tones, speech syllables) and long stimuli (piano melodies, speech sentences). In our natural

environment human observers are predominantly exposed to connected natural speech and piano music (e.g., melodies), thus, musicians should be familiar with the statistical structure of natural speech and piano music stimuli. Therefore, we expected that the effects of musical training would be more pronounced for long duration stimuli (melodies, speech sentences) as compared to short duration stimuli (piano music tones, speech syllables).

MATERIALS AND METHODS

PARTICIPANTS

Forty-one German native speakers gave informed consent to participate in the study (mean age \pm SD = 26 ± 4.9 years). Twenty-one subjects were amateur pianists (mean age \pm SD = 24.4 ± 5.1 years) with an average of 16.1 (SD = 5.3) years of experience of piano practicing (mean age of acquisition \pm SD = 8.2 ± 2.0 years), and they reported that they practiced the piano for an average of 3.48 (SD = 1.79) hours per week for the last 3 years. In the non-musicians group, all except three subjects (less than 3 months of music training in drums, bass guitar or flute) had no experience with practicing a musical instrument (mean age \pm SD = 27.8 ± 4.2 years). The study was approved by the joint human research review committee of the Max Planck Society and the University of Tübingen. A subset of these data (i.e., results for natural speech sentences and melodies) have previously been reported in Lee and Noppeney (2011a, 2014).

DESCRIPTION OF STIMULI

Synchronous audiovisual stimuli were recorded from one speaking actress uttering short sentences or one male hand playing on the piano keyboard (showing one octave) using a camcorder (HVX 200 P, Panasonic Corporation, Osaka, Japan; video at 25 frames per second, PAL 768*567 pixels) for the visual modality and analog recording for the auditory modality (2 channels, 48 kHz). The speech sentences were short neutral statements in German (4–5 words, 7–9 syllables). The music melodies were generated to match the rhythm and number of syllables to those of the speech sentences. The syllables were “do,” “re,” “mi,” “fa,” “so,” “la,” “di,” “to,” “bo,” “he,” “zi,” “ka,” “lo,” “ga,” “fi,” “po.” The piano music tones were “do,” “re,” “mi,” “fa,” “so,” “la,” “te,” “to.” Supplementary Material shows the list of speech sentences used in the experiment.

The visual and audio recordings were then digitized into MPEG-4 (H.264) format files. The visual file was first cropped to one single complete visual stimulus (speech or music), preceded and followed by 15 frames of neutral facial expression or a still hand image using Adobe Premier Pro (Adobe Systems, San Jose, CA, USA). We added the additional still images to be able to manipulated audiovisual asynchrony without changing the AV length of the stimuli (please see below and Maier et al., 2011).

To transform the auditory modality of natural speech into sinewave speech, the audio tracks were separated from the video tracks. The auditory natural speech was transformed into sinewave speech by replacing the three formants with sinusoid complexes of three sinusoids that were based on the first three vowel formants (www.lifesci.sussex.ac.uk/home/Chris_Darwin/Praatscripts/SWS). The auditory tracks of sinewave speech were re-combined with the video tracks to create

audiovisual movies of sinewave speech. Four sets of stimuli (24 stimuli per set; 8 stimuli per stimulus class) were created; two sets were stimuli of short duration (i.e., syllables or piano tones) and the other two sets were stimuli of long duration (i.e., sentences or melodies). The sets were counter balanced in time across subjects and across groups.

EXPERIMENTAL DESIGN

The experimental paradigm manipulated: (1) stimulus class: audiovisual speech, sinewave analogs of speech with visual utterance movements of natural speech, audiovisual piano music (i.e., piano music with associated hand movements), (2) stimulus duration: short (single syllables and single piano tones; mean duration $\pm SD = 2.38 \pm 0.37$ s; please note that the duration also include the 15 frame of still images before and after the action sequence), long (sentences and piano melodies; mean duration $\pm SD = 3.56 \pm 0.34$ s), and (3) audiovisual stimulus onset asynchronies (AV-SOA; ± 360 , ± 300 , ± 240 , ± 180 , ± 120 , ± 60 , 0 ms). Positive values indicated that the visual modality was presented first, whereas negative values indicated that the auditory modality was presented first. More specifically, in synchronous stimuli the temporal relationship between the video and the sound track was kept as obtained from recording and thus reflected the natural audiovisual temporal relationship. In other words, it complied with the natural statistics of audiovisual speech or music. Audiovisual asynchronous stimuli were generated by temporally shifting the onset of the auditory track with respect to the video. Moreover, audiovisual synchrony or asynchrony was then determined by the onset of the facial movements and sound rather than the onset of the video (for similar approach and rationale see Maier et al., 2011).

On each trial, subjects judged whether the audiovisual stimuli were synchronous or asynchronous, in an un-speeded fashion. They completed 8 sessions on 2 separate days. Each stimulus was presented 4 times per session in a randomized manner amounting to 2496 ($= 4$ sessions for each stimulus duration $\times 4$ times for each stimulus $\times 8$ stimuli per stimulus class $\times 3$ stimulus classes $\times 13$ AV-SOA) trials. The AV-SOA and stimulus class were randomized in each experiment. The stimuli of short and long duration were presented in separate sessions, and the order was counterbalanced across subjects and days. Prior to the experiment, subjects were presented with all stimuli (2 presentations per stimulus), and then tested on their comprehension of the SWS speech sentences by writing down each sentence that they hear.

EXPERIMENTAL PROCEDURE

The AV-SOA of the separate audio and video files was manipulated using Psychophysics Toolbox version 3 (PTB-3) under Matlab 2007b (MathWorks Inc., MA, USA). Visual stimuli (size $8.89^\circ \times 7^\circ$ visual angle) were projected using a CRT monitor (Sony Trinitron, Tokyo, Japan) at refresh rate of 100 Hz, and subjects' heads were stabilized using a chin rest. Auditory stimuli were presented at ~ 75 dB SPL via headphones.

DATA ANALYSIS

For each subject and condition, the proportion of synchronous responses (PSR) was computed for each of the 13 AV-SOA

levels. To refrain from making any distributional assumptions, the psychometric function was estimated using a non-parametric approach based on local linear fitting methods (Zychaluk and Foster, 2009). The bandwidth for the local quadratic fitting was optimized individually for each subject in a cross-validation procedure. We characterized the psychometric functions by the width of the temporal integration window, as determined by the integral of the psychometric function between -360 and $+360$ ms (after subtracting the difference between one and the maximum from all values of the fitted psychometric function, so that the maximum of all functions was set to one).

To evaluate whether there are any differences in the widths of the temporal integration window between groups, stimulus duration and stimulus class, mixed design ANOVAs were performed with stimulus duration (short, long) and stimulus class (natural speech, sinewave speech, piano music) as within-subject factors, and group (non-musicians, musicians) as a between-subject factor. The results of the ANOVAs are reported after Greenhouse-Geisser correction (when applicable).

RESULTS

After presenting subjects with all stimuli twice (before the main study), we tested them on the comprehension of sinewave speech sentences and syllables. Participants obtained 100% accuracy before the start of the experiment. This ensured that the intelligibility of sinewave speech stimuli could be considered speech-like for the main experiment.

Subjects' PSR for each condition was computed, and psychometric functions were estimated using a non-parametric local quadratic fitting method (Zychaluk and Foster, 2009). **Figure 1** shows the psychometric functions (averaged across subjects) separately for each condition in the musician and non-musician groups. **Figure 2** shows the bar plots of the mean (across subjects' mean) widths of the temporal integration windows for each condition in the musician and non-musician groups. The 2 (group: non-musicians, musicians) $\times 2$ (stimulus duration: short, long) $\times 3$ (natural speech, sinewave speech, piano music) mixed design ANOVA on the widths of the temporal integration windows (**Table 1**) revealed a main effect of stimulus duration. Thus, as previously suggested, participants accumulate information over time and thereby obtain more precise temporal estimates for long (i.e., melodies or sentences) relative to short duration stimuli (i.e., piano music tones or syllables) (Maier et al., 2011). Another previous study has reported the opposite finding, i.e., smaller temporal integration windows for syllables as compared to sentences (Vatakis and Spence, 2006a). Vatakis and Spence (2006a) have attributed their results to increased low-level spatiotemporal correlations or increased likelihood of binding attributable to the assumption of "unity" for long relative to short stimuli. However, this previous study differs from the current study in many aspects: (i) they used a temporal order judgment task, (ii) they included only very few stimuli (e.g., only two particular sentences), which makes generalization and interpretation difficult, and (iii) they investigated syllables and sentences in distinct sets of subjects and did not report a formal statistical comparison. For further discussion regarding the issue of stimulus duration, please refer to our previous study that aimed

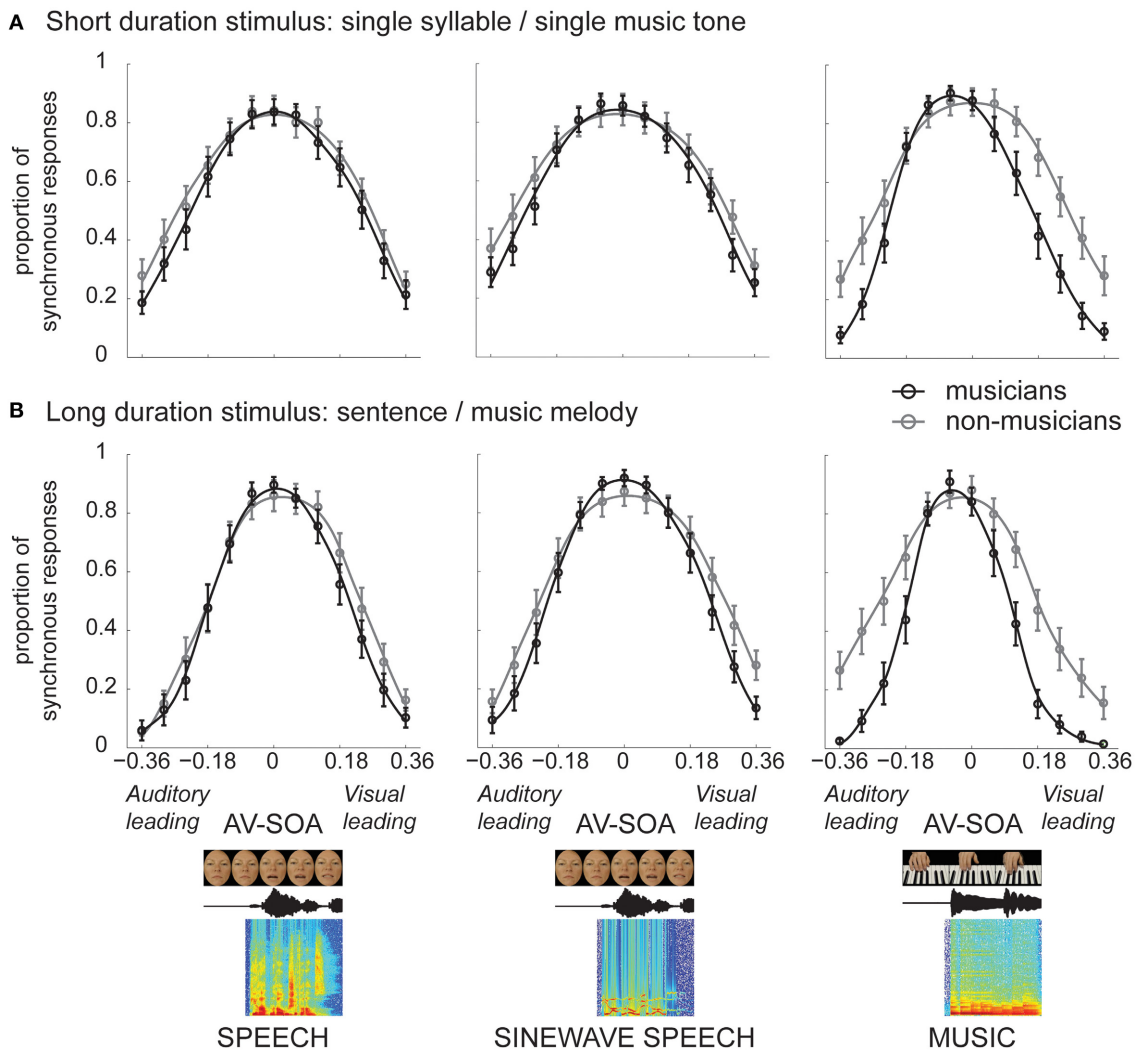


FIGURE 1 | The psychometric functions for speech, sinewave speech, and piano music in non-musicians and musicians for (A) short duration stimulus (syllables or single music tones), and (B) long duration stimulus (sentences or melodies).

to address the influence of stimulus duration on audiovisual temporal integration window (Maier et al., 2011).

Critically, we also observed main effects of stimulus class and group, as well as an interaction between stimulus class and group. As expected, music practice influenced musicians' temporal integration window in a stimulus-dependent fashion and had the strongest effect on piano music stimuli. Thus, as shown in **Figure 3**, the difference in widths of the temporal integration windows for musicians and non-musicians (i.e., the musical expertise effect) was the largest for piano music stimuli. Contrary to our initial hypothesis, we did not observe a significant three-way interaction of stimulus duration, stimulus class and group, a two-way interaction between group and duration, or a two-way interaction between stimulus class and duration. Therefore, we pooled the widths of the temporal integration windows across stimulus duration for natural speech, sinewave speech and piano music, and examined the effect of musical expertise for each

stimulus class by computing the difference of the mean widths of the temporal integration windows for musicians relative to non-musicians (i.e., musicians – non-musicians). **Figure 3** depicts the bar plots for the difference (musicians – non-musicians) of the mean widths of temporal integration windows for natural speech, sinewave speech and piano music. Specifically, we tested whether the musical expertise effect (i.e., the difference for musicians – non-musicians) on the widths of temporal integration windows was significantly greater than zero. *Post-hoc* two samples *t*-tests (one-tailed) for each stimulus class revealed that musicians relative to non-musicians exhibited significantly narrower temporal integration windows for sinewave speech [$t_{(39)} = 2.34$, $p = 0.025$; *one-tailed* $p = 0.01$] and piano music [$t_{(39)} = 4.74$, $p < 0.001$], and a marginal significance for natural speech [$t_{(39)} = 1.49$, $p = 0.14$; *one-tailed* $p = 0.07$]. Further, as illustrated in **Figure 3**, we observed a gradient of musical expertise effects for piano music > sinewave speech > natural speech. This

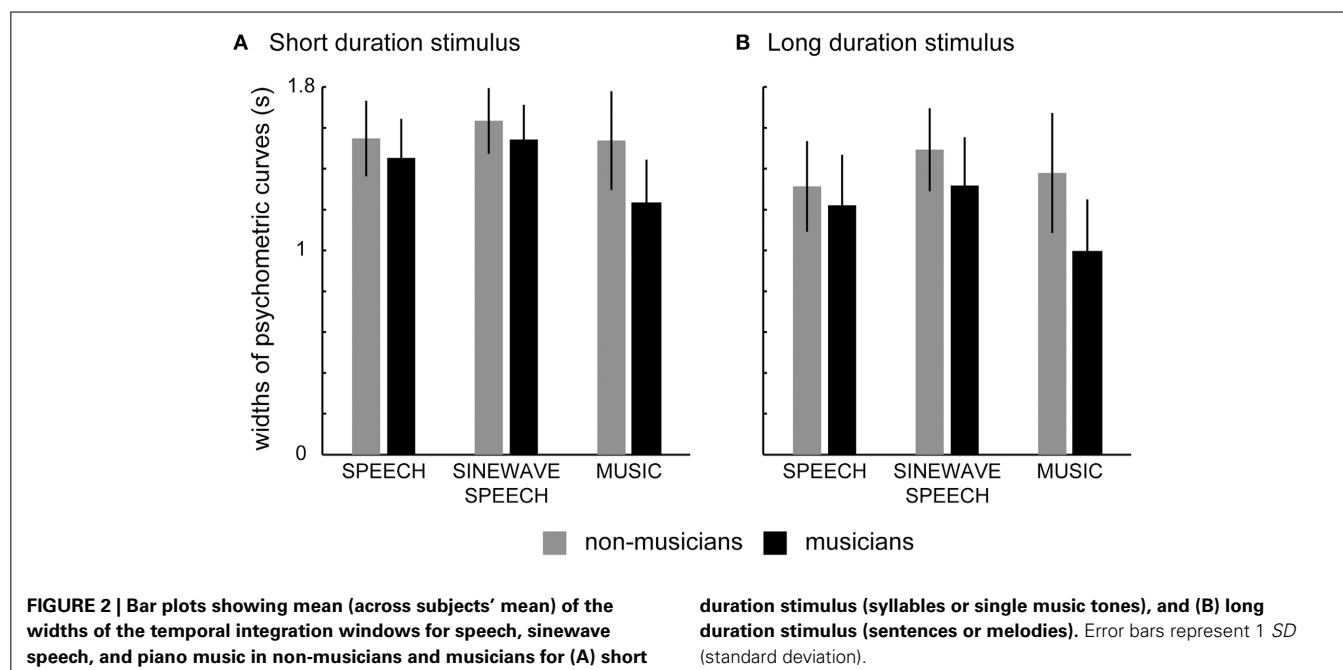


Table 1 | Results of the mixed ANOVA on the widths of temporal integration windows with stimulus duration (short, long) and stimulus class (speech, sinewave speech, music) as within-subject factors, and group (non-musicians, musicians) as between-subject factor.

Main effects of:		
Group	$F_{(1, 39)} = 10.08$	$p = 0.003$
Stimulus duration	$F_{(1, 39)} = 129.5$	$p < 0.001$
Stimulus class	$F_{(1.48, 57.8)} = 53.5$	$p < 0.001$
Interactions of:		
Group * stimulus duration	$F_{(1, 39)} = 2.20$	$p = 0.146$
Group * stimulus class	$F_{(1.48, 57.8)} = 22.0$	$p < 0.001$
Stimulus duration * stimulus class	$F_{(1.46, 56.9)} = 1.60$	$p = 0.215$
Group * stimulus duration * stimulus class	$F_{(1.46, 56.9)} = 1.44$	$p = 0.243$

Significant effects are indicated in bold.

observation was confirmed statistically by *post-hoc* testing for the three interactions that selectively compare the musical expertise effect across two stimulus classes (e.g., musicians – non-musicians for piano music – natural speech). These tests demonstrated that musicians relative to non-musicians exhibited narrower temporal integration windows for piano music > natural speech [$t_{(39)} = 5.41, p < 0.001$] and music > sinewave speech [$t_{(39)} = 4.58, p < 0.001$], and a marginal significance for sinewave speech > natural speech [$t_{(39)} = 1.51, p = 0.14$; *one-tailed* $p = 0.07$]. A one-tailed *t*-test can be adopted, because we would expect a stronger musical expertise effect for sinewave speech than natural speech stimuli (see Introduction).

Collectively, these results demonstrate that the effect of musical expertise was most pronounced for piano music stimuli, and it also generalized to sinewave speech and to a marginally significant

extent to natural speech stimuli. However, contrary to our expectations, the musical expertise effect did not depend on stimulus duration. This suggests that even short stimuli provided sufficient statistical structure that enabled musicians to generate more precise estimates of the relative timing of the audiovisual signals.

CORRELATION ANALYSES OF THE WIDTHS OF AUDIOVISUAL TEMPORAL INTEGRATION WINDOWS WITH AGE OF ACQUISITION AND AMOUNT OF PRACTICE

The narrowing of the temporal integration window for musicians may result from innately specified (e.g., genetic) differences between musicians and non-musicians. Alternatively, it may reflect plasticity induced by long-term musical training (Munte, 2002; Zatorre et al., 2007). In the latter case, the narrowing of the temporal integration may depend on the amount of time that musicians spent on piano practicing. Further, the effect of music practice may also interact with neurodevelopment and be most pronounced when children start practicing a musical instrument early in life. In this case, the effect of music practicing should depend on the age at which musicians started piano practicing. Effects of age of acquisition would for instance be observed if piano practicing relies on mechanisms that need to be fine-tuned during sensitive periods in neurodevelopment.

To test whether the narrowing of temporal integration window results from training-induced plasticity, we performed separate correlation analyses testing for a correlation between the width of the psychometric function with (i) age of acquisition or (ii) amount of weekly music practice (in hours) during the past 3 years as predictors. As the widths of temporal integration windows were highly correlated across the different conditions over subjects, we first performed a principal component analysis on the subject-specific widths across all conditions for data reduction. The first component explained 76.3% of the total variance

of all the widths, while the second component explained 9.1% and the third component explained 6.2% of the total variance of all the widths. Thus, as the 2nd component explained only a negligible amount of variance in the data, we extracted and

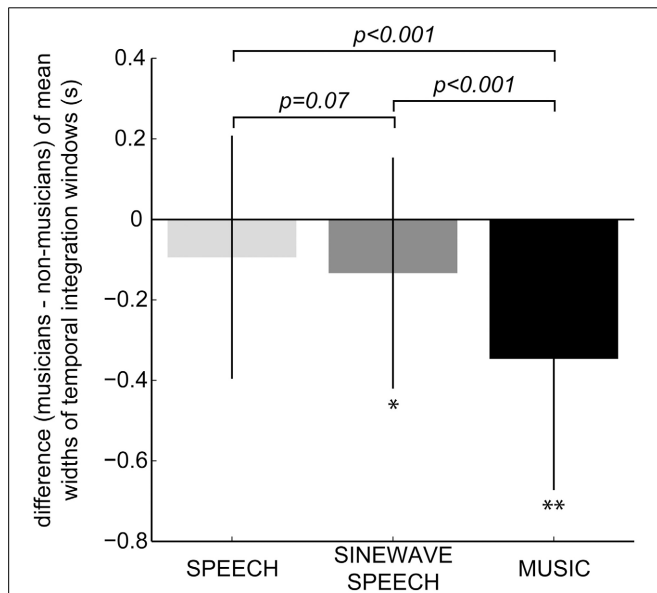


FIGURE 3 | Bar plots showing the musical expertise effect, i.e., difference of the mean widths of the temporal integration windows for musicians vs. non-musicians (musicians – non-musicians; a negative value indicated that musicians relative to non-musicians exhibited a narrower temporal integration window). Error bars represent 1 SD (standard deviation). Significance was calculated using one-tailed two samples *t*-tests on the difference of the mean widths of the temporal integration windows for musicians vs. non-musicians (* $p < 0.05$; ** $p < 0.001$). Additionally, *p*-values (one-tailed) of one-tailed two samples *t*-tests of the musical expertise effects for sinewave speech > natural speech, piano music > sinewave speech, and piano music > natural speech are shown.

correlated only the first component with age of acquisition and amount of weekly piano music practice during the past 3 years. A significant correlation was found for the first component and amount of weekly piano music practice during the past 3 years [$r_{(21)} = -0.46$, $p = 0.037$] (**Figure 4A**), whereas no significant correlation was found for the first component and age of acquisition [$r_{(21)} = 0.116$, $p = 0.617$] (**Figure 4B**). Specifically, the more the musicians practiced piano, the narrower their temporal integration windows were (i.e., the more sensitive they became to the temporal misalignment of auditory and visual signals).

DISCUSSION

Our results demonstrate that long-term music training shapes the temporal integration window in a stimulus-dependent fashion. Musicians, relative to non-musicians, exhibited a narrower temporal integration window predominantly for piano music and to some extent also for sinewave speech with a marginally significant trend for natural speech. Moreover, the amount of weekly piano music practice in the past 3 years correlated with the widths of the temporal integration windows across all stimulus classes. In other words, the more musicians practiced piano in the past 3 years, the more sensitive they became to audiovisual temporal misalignments for natural speech, sinewave speech, and music. Collectively, our results demonstrate that music practice furnishes more precise estimates regarding the relative timings of the audiovisual signals predominantly for music, yet this effect also transferred partly to speech.

Accumulating evidence suggests that music practicing and perceptual learning can influence how human observers temporally bind signals from multiple senses. For instance, a recent psychophysics study demonstrated that musical expertise narrows the temporal integration window for music (Petrini et al., 2009). Yet, this study included only music stimuli. Thus, an unresolved question is to what extent these music or perceptual learning effects are specific to the particular stimulus class trained or whether they can generalize to other stimulus classes. In support

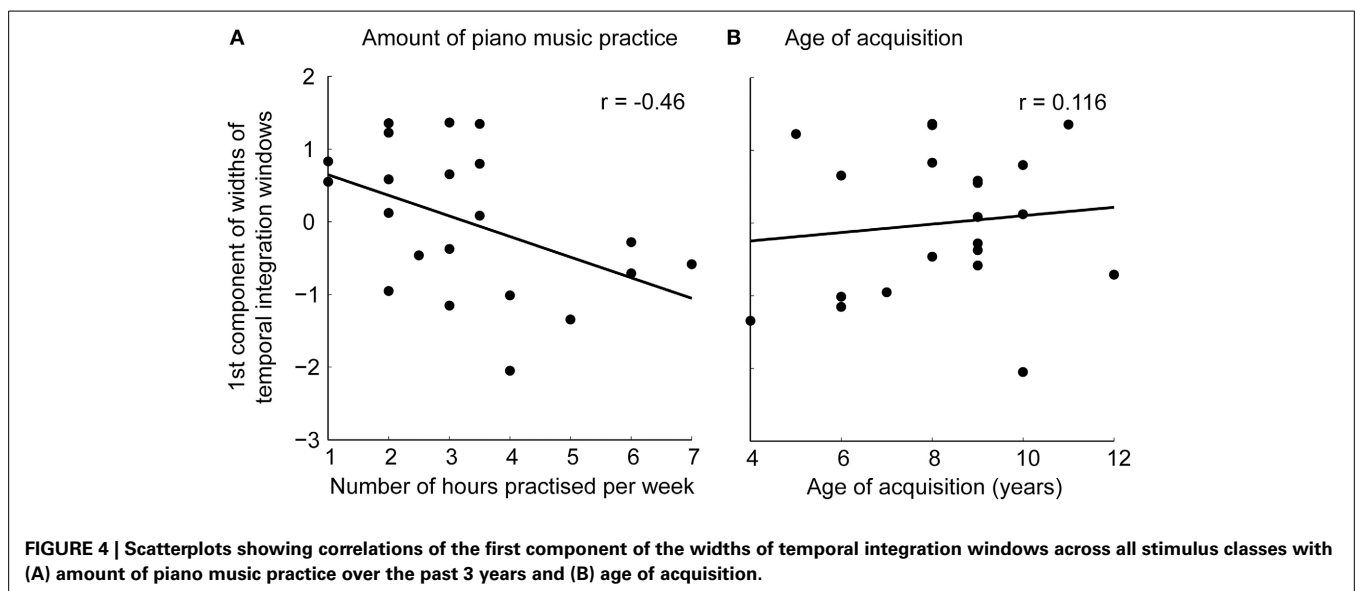


FIGURE 4 | Scatterplots showing correlations of the first component of the widths of temporal integration windows across all stimulus classes with (A) amount of piano music practice over the past 3 years and (B) age of acquisition.

of generic mechanisms of musical expertise, electrophysiological recording demonstrated earlier and larger brain stem responses for musicians relative to non-musicians for both speech and music (Musacchia et al., 2007, 2008; Bidelman and Krishnan, 2010; Bidelman et al., 2011). By contrast, a recent neuroimaging study demonstrated that music practice fine-tunes the temporal integration window predominantly for piano music via engagement of a premotor-cerebellar circuitry (Lee and Noppeney, 2011a).

The current study therefore revisited the question of whether music practice influences audiovisual temporal integration not only of the trained piano music stimuli but also untrained stimulus classes. To this aim, we included natural speech and intelligible sinewave speech signals where the main speech formants have been replaced by sinewave analogs, thereby giving sinewave speech a musical character. Critically, even though the sinewave speech transformation preserved stimulus intelligibility, it introduced a novel mapping between auditory and visual signals. Indeed, as expected, this novel audiovisual mapping made it harder for participants to discriminate between synchronous and asynchronous audiovisual sinewave speech as indicated by a broader integration window for sinewave speech as compared to natural speech (for related findings on rotated speech, see Maier et al., 2011). Thus, the comparison of piano music, sinewave speech and natural speech stimuli enabled us to better characterize to which extent music practice effects transfer to other stimulus classes.

Our results replicate that music expertise shapes temporal binding of audiovisual signals in a stimulus-dependent fashion as indicated by a significant interaction between stimulus class and group. Thus, we observed a gradient of musical expertise effects decreasing from piano music > sinewave speech > natural speech. Nevertheless, the effects of musical expertise on the temporal integration window of other stimulus classes such as sinewave speech or natural speech were still significant. The gradient of musical expertise effect across stimulus classes may be accounted for by two different explanatory frameworks:

First, audiovisual temporal perception may be mediated by only one domain-general mechanism that is engaged by all stimulus classes. Since this domain-general system can be fine-tuned via training to the statistics of a particular stimulus class, the musical expertise effect varies across stimulus classes in a gradual fashion. Thus, pianists would be particularly sensitive to audiovisual asynchronies of piano music stimuli, because the domain-general system has been fine-tuned to the audiovisual temporal statistics of piano music. Yet, transfer effects of musical expertise also emerge, because other stimulus classes can benefit from the fine-tuning of a domain-general system.

Alternatively, the gradient in musical expertise effects may be explained by the concurrent engagement of domain-general and stimulus-specific mechanisms. Domain-general mechanisms have been proposed by a vast number of studies showing musical expertise effects that generalize across music and speech stimuli at the behavioral (Chandrasekaran et al., 2009; Elmer et al., 2012, 2013; Marie et al., 2012; Asaridou and McQueen, 2013) or neural level (Musacchia et al., 2008; Bidelman and Krishnan, 2010; Bidelman et al., 2011; Elmer et al., 2012, 2013; Marie et al.,

2012). Conversely, we recently showed that music practice sharpens the temporal integration window predominantly for music via premotor-cerebellar circuitry (Lee and Noppeney, 2011a) and proposed that piano practicing may mold the audiovisual temporal integration by training an internal forward model that maps from motor actions (e.g., piano practicing) to its sensory consequences in vision (i.e., finger movements) and audition (e.g., piano sound when hitting the key). Thus, a combination of such a domain-general and a stimulus-dependent sensory-motor mechanism may better explain the transfer of musical expertise effects to other stimulus classes such as sinewave speech in a gradual fashion.

The comparison of musicians and non-musicians cannot resolve ambiguities about whether or not the mechanisms are innately specified or truly reflect experience-dependent plasticity. For instance, amateur musicians may have chosen to practice a musical instrument, because they were inherently better at temporal perception via innate mechanisms. Yet, if musical expertise depends on experience-dependent mechanisms, we would expect that the temporal integration window decrease with the amount of practice. Moreover, if these experience-dependent mechanisms interact with development (e.g., sensitive periods), the integration window should also be influenced by the age at which participants started practicing a musical instrument. Our results demonstrate that indeed the amount of weekly piano practicing in the past 3 years correlates negatively with the musicians' widths of the temporal integration windows—more specifically the first principal component over widths across all conditions. In other words, the more musicians practiced piano, the more sensitive they were to audiovisual temporal misalignments of speech and piano music stimuli. Surprisingly, the age at which musicians started piano practicing did not correlate significantly with the widths of their temporal integration windows. This dissociation suggests that piano practicing shapes audiovisual temporal integration and sensitivity to temporal misalignments via experience-dependent mechanisms that either do not critically interact with neurodevelopment or are bound to sensitive periods in very early development (i.e., before the age of four when the first of our participants started piano practicing). Yet, our results are based on correlative methods. To further substantiate our conclusions, prospective longitudinal studies are required that investigate the change in the temporal integration window as a function of piano music practicing [e.g., 2 (piano practicing vs. other activity) \times 2 (before, after training) factorial design].

In conclusion, our results suggest that piano music practicing shapes the temporal integration of audiovisual signals via experience-dependent plasticity. While musical expertise strongly narrows the width of the temporal integration window for piano music, the effect transfers to non-music stimuli such as sinewave speech and a non-significant trend to natural speech. Thus, piano music practicing affects temporal binding either via mechanisms that are specialized predominantly for music but transfer at least in part to other stimulus classes. Alternatively, piano music practicing influences temporal binding of audiovisual signals via multiple mechanisms including stimulus-specific (i.e., specialized for music stimuli) and generic mechanisms (e.g., perceptual learning).

ACKNOWLEDGMENTS

This work is funded by Max Planck Society. We thank Fabian Sinz for creating the music stimuli, Mario Kleiner for help to program the experiment using Matlab with PsychToolbox extensions, Prof. K. Zychaluk and Prof. D. H. Foster for useful discussion regarding the fitting and characterization of the psychometric functions, and Dr. Fabrizio Leo and Dr. Massimiliano Di Luca for comments on the earlier drafts of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fpsyg.2014.00868/abstract>

REFERENCES

- Alais, D., and Burr, D. (2003). The “flash-lag” effect occurs in audition and cross-modally. *Curr. Biol.* 13, 59–63. doi: 10.1016/S0960-9822(02)01402-1
- Asaridou, S. S., and McQueen, J. M. (2013). Speech and music shape the listening brain: evidence for shared domain-general mechanisms. *Front. Psychol.* 4:321. doi: 10.3389/fpsyg.2013.00321
- Baart, M., Stekelenburg, J. J., and Vroomen, J. (2014). Electrophysiological evidence for speech-specific audiovisual integration. *Neuropsychologia* 53, 115–121. doi: 10.1016/j.neuropsychologia.2013.11.011
- Baumann, S., Meyer, M., and Jancke, L. (2008). Enhancement of auditory-evoked potentials in musicians reflects an influence of expertise but not selective attention. *J. Cogn. Neurosci.* 20, 2238–2249. doi: 10.1162/jocn.2008.20157
- Besson, M., Schon, D., Moreno, S., Santos, A., and Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restor. Neurol. Neurosci.* 25, 399–410.
- Bidelman, G. M., and Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Res.* 1355, 112–125. doi: 10.1016/j.brainres.2010.07.100
- Bidelman, G. M., Krishnan, A., and Gandour, J. T. (2011). Enhanced brainstem encoding predicts musicians’ perceptual advantages with pitch. *Eur. J. Neurosci.* 33, 530–538. doi: 10.1111/j.1460-9568.2010.07527.x
- Chandrasekaran, B., Krishnan, A., and Gandour, J. T. (2009). Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain Lang.* 108, 1–9. doi: 10.1016/j.bandl.2008.02.001
- Dixon, N. F., and Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception* 9, 719–721.
- Elmer, S., Hanggi, J., Meyer, M., and Jancke, L. (2013). Increased cortical surface area of the left planum temporale in musicians facilitates the categorization of phonetic and temporal speech sounds. *Cortex* 49, 2812–2821. doi: 10.1016/j.cortex.2013.03.007
- Elmer, S., Meyer, M., and Jancke, L. (2012). Neurofunctional and behavioral correlates of phonetic and temporal categorization in musically trained and untrained subjects. *Cereb. Cortex* 22, 650–658. doi: 10.1093/cercor/bhr142
- Fujisaki, W., Shimojo, S., Kashino, M., and Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nat. Neurosci.* 7, 773–778. doi: 10.1038/nn1268
- Grant, K. W., Van Wassenhove, V., and Poeppel, D. (2004). Detection of auditory (cross-spectral) and auditory-visual (cross-modal) synchrony. *Speech Commun.* 44, 43–53. doi: 10.1016/j.specom.2004.06.004
- Hannon, E. E., and Trainor, L. J. (2007). Music acquisition: effects of enculturation and formal training on development. *Trends Cogn. Sci.* 11, 466–472. doi: 10.1016/j.tics.2007.08.008
- Imfeld, A., Oechslin, M. S., Meyer, M., Loenneker, T., and Jancke, L. (2009). White matter plasticity in the corticospinal tract of musicians: a diffusion tensor imaging study. *Neuroimage* 46, 600–607. doi: 10.1016/j.neuroimage.2009.02.025
- Kraus, N., and Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nat. Rev. Neurosci.* 11, 599–605. doi: 10.1038/nrn2882
- Lee, H., and Noppeney, U. (2011a). Long-term music training tunes how the brain temporally binds signals from multiple senses. *Proc. Natl. Acad. Sci. U.S.A.* 108, E1441–E1450. doi: 10.1073/pnas.1115267108
- Lee, H., and Noppeney, U. (2011b). Physical and perceptual factors shape the neural mechanisms that integrate audiovisual signals in speech comprehension. *J. Neurosci.* 31, 11338–11350. doi: 10.1523/JNEUROSCI.6510-10.2011
- Lee, H., and Noppeney, U. (2014). Temporal prediction errors in visual and auditory cortices. *Curr. Biol.* 24, R309–R310. doi: 10.1016/j.cub.2014.02.007
- Lewis, R., and Noppeney, U. (2010). Audiovisual synchrony improves motion discrimination via enhanced connectivity between early visual and auditory areas. *J. Neurosci.* 30, 12329–12339. doi: 10.1523/JNEUROSCI.5745-09.2010
- Love, S. A., Petrini, K., Cheng, A., and Pollick, F. E. (2013). A psychophysical investigation of differences between synchrony and temporal order judgments. *PLoS ONE* 8:e54798. doi: 10.1371/journal.pone.0054798
- Magne, C., Schon, D., and Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. *J. Cogn. Neurosci.* 18, 199–211. doi: 10.1162/089892906775783660
- Maier, J. X., Di Luca, M., and Noppeney, U. (2011). Audiovisual asynchrony detection in human speech. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 245–256. doi: 10.1037/a0019952
- Marie, C., Kujala, T., and Besson, M. (2012). Musical and linguistic expertise influence pre-attentive and attentive processing of non-speech sounds. *Cortex* 48, 447–457. doi: 10.1016/j.cortex.2010.11.006
- Marques, C., Moreno, S., Castro, S. L., and Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: behavioral and electrophysiological evidence. *J. Cogn. Neurosci.* 19, 1453–1463. doi: 10.1162/jocn.2007.19.9.1453
- Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., and Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. *Cereb. Cortex* 19, 712–723. doi: 10.1093/cercor/bhn120
- Munte, T. F. (2002). Brains out of tune. *Nature* 415, 589–590. doi: 10.1038/415589a
- Munte, T. F., Altenmüller, E., and Jancke, L. (2002). The musician’s brain as a model of neuroplasticity. *Nat. Rev. Neurosci.* 3, 473–478. doi: 10.1038/nrn843
- Musacchia, G., Sams, M., Skoe, E., and Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15894–15898. doi: 10.1073/pnas.0701498104
- Musacchia, G., Strait, D., and Kraus, N. (2008). Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hear. Res.* 241, 34–42. doi: 10.1016/j.heares.2008.04.013
- Navarra, J., Vatakis, A., Zampini, M., Soto-Faraco, S., Humphreys, W., and Spence, C. (2005). Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. *Brain Res. Cogn. Brain Res.* 25, 499–507. doi: 10.1016/j.cogbrainres.2005.07.009
- Noesselt, T., Bonath, B., Boehler, C. N., Schoenfeld, M. A., and Heinze, H. J. (2008). On perceived synchrony—neural dynamics of audiovisual illusions and suppressions. *Brain Res.* 1220, 132–141. doi: 10.1016/j.brainres.2007.09.045
- Noesselt, T., Rieger, J. W., Schoenfeld, M. A., Kanowski, M., Hinrichs, H., Heinze, H. J., et al. (2007). Audiovisual temporal correspondence modulates human multisensory superior sulcus plus primary sensory cortices. *J. Neurosci.* 27, 11431–11441. doi: 10.1523/JNEUROSCI.2252-07.2007
- Parbery-Clark, A., Strait, D. L., Hittner, E., and Kraus, N. (2013). Musical training enhances neural processing of binaural sounds. *J. Neurosci.* 33, 16741–16747. doi: 10.1523/JNEUROSCI.5700-12.2013
- Petrini, K., Dahl, S., Rocchesso, D., Waadeland, C. H., Avanzini, F., Puce, A., et al. (2009). Multisensory integration of drumming actions: musical expertise affects perceived audiovisual asynchrony. *Exp. Brain Res.* 198, 339–352. doi: 10.1007/s00221-009-1817-2
- Powers, A. C. 3rd., Hillock, A. R., and Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *J. Neurosci.* 29, 12265–12274. doi: 10.1523/JNEUROSCI.3501-09.2009
- Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science* 212, 947–949.
- Schlaug, G., Jancke, L., Huang, Y., and Steinmetz, H. (1995). *In vivo* evidence of structural brain asymmetry in musicians. *Science* 267, 699–701.
- Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H. J., et al. (2005). Structural and functional asymmetry of lateral Heschl’s gyrus reflects pitch perception preference. *Nat. Neurosci.* 8, 1241–1247. doi: 10.1038/nn1530
- Spence, C., and Squire, S. (2003). Multisensory integration: maintaining the perception of synchrony. *Curr. Biol.* 13, R519–R521. doi: 10.1016/S0960-9822(03)00445-7

- Stein, B. E., Meredith, M. A., and Wallace, M. T. (1993). The visually responsive neuron and beyond: multisensory integration in cat and monkey. *Prog. Brain Res.* 95, 79–90.
- Stekelenburg, J. J., and Vroomen, J. (2012). Electrophysiological evidence for a multisensory speech-specific mode of perception. *Neuropsychologia* 50, 1425–1431. doi: 10.1016/j.neuropsychologia.2012.02.027
- Stevenson, R. A., Vanderklok, R. M., Pisoni, D. B., and James, T. W. (2011). Discrete neural substrates underlie complementary audiovisual speech integration processes. *Neuroimage* 55, 1339–1345. doi: 10.1016/j.neuroimage.2010.12.063
- Thompson, W. F., Schellenberg, E. G., and Husain, G. (2004). Decoding speech prosody: do music lessons help? *Emotion* 4, 46–64. doi: 10.1037/1528-3542.4.1.46
- Tzounopoulos, T., and Kraus, N. (2009). Learning to encode timing: mechanisms of plasticity in the auditory brainstem. *Neuron* 62, 463–469. doi: 10.1016/j.neuron.2009.05.002
- van Wassenhove, V., Grant, K. W., and Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia* 45, 598–607. doi: 10.1016/j.neuropsychologia.2006.01.001
- Vatakis, A., and Spence, C. (2006a). Audiovisual synchrony perception for music, speech, and object actions. *Brain Res.* 1111, 134–142. doi: 10.1016/j.brainres.2006.05.078
- Vatakis, A., and Spence, C. (2006b). Audiovisual synchrony perception for speech and music assessed using a temporal order judgment task. *Neurosci. Lett.* 393, 40–44. doi: 10.1016/j.neulet.2005.09.032
- Vatakis, A., and Spence, C. (2007). Crossmodal binding: evaluating the “unity assumption” using audiovisual speech stimuli. *Percept. Psychophys.* 69, 744–756. doi: 10.3758/BF03193776
- Vatakis, A., and Spence, C. (2008a). Evaluating the influence of the ‘unity assumption’ on the temporal perception of realistic audiovisual stimuli. *Acta Psychol.* 127, 12–23. doi: 10.1016/j.actpsy.2006.12.002
- Vatakis, A., and Spence, C. (2008b). Investigating the effects of inversion on configural processing with an audiovisual temporal-order judgment task. *Perception* 37, 143–160. doi: 10.1068/p5648
- Vroomen, J., and Stekelenburg, J. J. (2011). Perception of intersensory synchrony in audiovisual speech: not that special. *Cognition* 118, 75–83. doi: 10.1016/j.cognition.2010.10.002
- Zampini, M., Guest, S., Shore, D. I., and Spence, C. (2005). Audio-visual simultaneity judgments. *Percept. Psychophys.* 67, 531–544. doi: 10.3758/BF03193329
- Zatorre, R. J., Chen, J. L., and Penhune, V. B. (2007). When the brain plays music: auditory-motor interactions in music perception and production. *Nat. Rev. Neurosci.* 8, 547–558. doi: 10.1038/nrn2152
- Zychaluk, K., and Foster, D. H. (2009). Model-free estimation of the psychometric function. *Atten. Percept. Psychophys.* 71, 1414–1425. doi: 10.3758/APP.71.6.1414

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 01 April 2014; accepted: 21 July 2014; published online: 07 August 2014.

Citation: Lee H and Noppeney U (2014) Music expertise shapes audiovisual temporal integration windows for speech, sinewave speech, and music. *Front. Psychol.* 5:868. doi: 10.3389/fpsyg.2014.00868

This article was submitted to Language Sciences, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Lee and Noppeney. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.